DOCUMENT RESUME

ED 327 584 TM 016 039

AUTHOR Jibaja-Rusth, Maria; And Others
TITLE Measurement Characteristics of the Stages of Concerns Questionnaire (SoCQ) during Baseline Phases.
PUB DATE 17 Jan 91
NOTE 14p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 24, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Curriculum Development; Longitudinal Studies; Measurement; Models; Pilot Projects; *Pretests Posttests; Questionnaires; Science Education; *Science Teachers; Secondary Education; *Secondary School Teachers; *Test Reliability
IDENTIFIERS Alpha Coefficient; *Internal Consistency; *Stages of Concern Questionnaire

ABSTRACT

The purpose of this study was to investigate the internal consistency reliability of the Stages of Concern Questionnaire (SoCQ) when the measure is administered to naive subjects as a pretest. The SoCQ was developed by G. E. Hall, A. A. George, and W. L. Rutherford (1979/1986). In the present study, the SoCQ was administered to secondary school science teachers participating in the pilot of a curriculum model that: (1) exposed students to more hours of science over a longer period of time; (2) immersed students in a program reinforcing the interrelatedness of science disciplines; and (3) built from a descriptive knowledge base toward abstract thought processes. Data are reported for the first three measurement points in the study: point-0 (pretest) data were collected from 22 teachers before the implementation of the revised curriculum; point-1 data were collected from 15 teachers after about 1 week; and point-2 data were collected from 25 teachers, including 12 members of the original cohort, at about 6 months into the implementation of the intervention. The alpha coefficients from the various points gave marginal results, and were particularly unfavorable for the awareness scale. Results neither confirm the plausibility of the concerns model nor indicate that the SoCQ yielded reliable scores for all the model's constructs. A 21-item list of references and 1 table, which summarizes the alpha coefficients, are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED327584

lesliem4.wp1 1/17/91

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BRUCE THOMPSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MEASUREMENT CHARACTERISTICS OF THE STAGES OF CONCERNS QUESTIONNAIRE (SoCQ) DURING BASELINE PHASES

Maria Jibaja-Rusth
Judith H. Dresden
Linda W. Crow

Baylor College of Medicine

Bruce Thompson

Texas A&M University 77843-4225
and
Baylor College of Medicine

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January 24, 1991.

TM016039

ABSTRACT

Measurement reliability is critical to viable substantive inquiry, so it is important to explore the reliability of scores in hand in any study. Particular problems in acquiring reliable data are confronted when subjects are naive, as they may be before a needed intervention, or when scores are collected relative to stages in developmental models. The present paper explores the reliability of data collected from subjects using the Stages of Concern Questionnaire. Results suggest that the measure must be used with some caution.

Perhaps the most common misperception about measurement is the view that reliability is a characteristic that inures to tests. Instead, reliability is a characteristic of **data**, albeit data generated on a given measure administered with a given protocol to given subjects on given occasions. As Rowley (1976, p. 53) notes, "It needs to be established that an instrument itself is neither reliable nor unreliable." As Sax (1980, p. 261) explains,

> It is more accurate to talk about the reliability of measurements (data, scores, and observations) than the reliability of tests (questions, items, and other tasks). Tests cannot be stable or unstable, but observations can. Any reference to the "reliability of a test" should always be interpreted to mean the "reliability of measurements or observations [i.e., a particular set of data] derived from a test."

Similarly, Crocker and Algina (1986, p. 144, emphasis added) note that, "a test is not 'reliable' or 'unreliable.' Rather, reliability is a property of the **scores** on a test for a particular group of examinees."

For example, it is difficult to generalize the "reliability of a test", such as a twelfth-grade achievement measure, when the test is administered to fourth-graders. Theoretically, the fourth-graders would guess on most or all of the questions, the responses would be random, and the reliability of the data from the so-called "reliable" test would be zero.

A similar paradox confronts persons administering tests to naive subject groups for pretest purposes. In the cognitive domain, if subjects all know nothing about a given domain, and truly need the intervention, then their responses on the pretest will be random and perfectly unreliable. Similarly, in the attitude domain, subjects tend to not generate reliable data in areas where they lack preconceptions.

In reality, some subjects usually know or feel at least something about a measurement domain, and the paradox occurs in a less extreme form. Nevertheless, the researcher pretesting naive subjects confronts some unique obstacles, and cannot assume that measures that function well as posttests will also function well in pretest applications.

The difficulties posed in gathering reliable scores are dramatically exacerbated when one is collecting data on a hierarchy of constructs that subjects are expected in theory to work through in a developmental sequence. As the salience of constructs shift through the developmental process, the reliability of the scores gathered on the constructs should then vary at different points in time.

In any case, the realization that tests are not reliable implies, among other things, that researchers should generally empirically investigate the reliability of scores in hand, to confirm that reliable data have been collected. The results of such analyses can inform the researcher whether revised measurement is necessary, and also provide a conceptual framework for interpreting

obtained substantive effect sizes, given that measurement error attenuates the magnitudes of the effects that can be statistically detected (Reinhardt, 1991).

Purpose of the Paper

The purpose of the present paper is to investigate the internal consistency reliability of the Stages of Concern Questionnaire (SoCQ) when the measure is administered to naive subjects as a pretest. The SoCQ was developed by Hall, George and Rutherford (1979/1986). The theoretical underpinnings of the measure are described by Hall and his colleagues (cf. Hall, 1978; Hall & Hord, 1987; Hall, Loucks, Rutherford & Newlove, 1975; Hord, Rutherford, Huling-Austin & Hall, 1987; Loucks & Hall, 1977). The instrument has proven useful in a number of studies (Broyles & Tillman, 1985; Cicchelli & Baecher, 1987; Holloway & Kerr, 1979; James & Hall, 1981; Kimpston & Anderson, 1985; Wedman & Heller, 1984). Nevertheless, because the measure purportedly taps a developmental hierarchy, some of the problems mentioned previously may affect scores on the measure.

Hall et al. (1979/1986, p. 11) do present coefficient alpha for each of the seven SoCQ scales based on data from 830 naive subjects at the start of an intervention. The coefficients ranged from .64 to .83. These results illustrate the difficulty of securing reliable data from naive subjects.

Indeed, such coefficients may distort substantive findings to an unacceptable degree. As Locke, Spirduso and Silverman (1987, p. 28) note, "the correlation between scores from two tests cannot

exceed the square root of the product for reliability in each test." If the pretest alpha for the Awareness scale was .64 for data in hand, as it was in the Hall et al. (1979/1986) study, the mathematical ceiling on the correlation between Awareness scores and even a perfectly reliable second variable would be .80 ((.64 x 1.00) ^ .5).

Method

Subjects

For the purposes of the present study the SoCQ was administered to secondary school teachers from one of four schools in one of two school districts. The teachers were science teachers participating in a pilot of the Scope, Sequence, and Coordination (SSC) curriculum model sponsored by the National Science Teachers Association. This model (a) exposes students to more hours of science over a longer period of time, (b) immerses students in a program reinforcing the interrelatedness of science disciplines, and (c) builds from a descriptive knowledge base, through empirical study, then to abstract thought processes.

The data constituted the initial entries in a small longitudinal data base involving SoCQ scores of teachers participating in the intervention. Data reported here were collected at the first three measurement points in the study: (a) Point 0 pretest data were collected prior to the intervention's implementation from 22 teachers, including 15 teachers who were part of the longitudinal study; (b) Point 1 data were collected roughly one week later from the 15 teachers in the longitudinal

cohort; (c) Point 2 data were collected from 25 teachers from the same schools, including 12 members of the longitudinal cohort, roughly six months into the implementation of the intervention.

Analyses in the present study were performed both for all data available and for cohort members separately. Reliability coefficients tend to be attenuated by homogeneous samples (Reinhardt, 1991), so more favorable values were expected to emerge from analyses including teachers not in the longitudinal cohort.

Results

Table 1 presents the alpha coefficients computed in the present study across the three points in time. Table 1 also presents the related results reported by Hall et al. (1979), and by Bailey and Palsha (1991) in their study involving 142 subjects from three states.

INSERT TABLE 1 ABOUT HERE.

Discussion

As Neale and Liebert (1986, p. 290) observe:

> No one study, however shrewdly designed and carefully executed, can provide convincing support for a causal hypothesis or theoretical statement... Too many possible (if not plausible) confounds, limitations on generality, and alternative interpretations can be offered for any one observation.... How, then, does social science theory advance through research? The answer is, by

collecting a diverse body of evidence about any major theoretical proposition.

And certainly generalizing from the preliminary results of our small longitudinal study alone would be risky business.

But it is disturbing that the alpha coefficients reported in various studies at various points in the implementation of innovations and involving various sample sizes consistently report results that are somewhat marginal. And results for the Awareness scale are particularly unfavorable.

Bailey and Palsha (1991) suggest one intriguing possible explanation for such results--the concerns model may not be fully viable. They propose an alternative five-factor model that they found fit their data better than the seven-factor SoCQ model.

Bailey and Palsha (1991) avoided one common misconception seen all too often in published research, i.e., the misconception that a large alpha coefficient implies a unidimensional factor structure (Crocker & Algina, 1986, pp. 142-143). While factors in a viable model should yield reliable construct scores, the obverse is not inherently true--measures yielding scores with high internal consistency may quite often be factored into various subdimensions.

Additional revision of the SoCQ does appear to be warranted. While many researchers have found the conceptual model to be useful, as noted previously, it is not yet clear that the plausibility of the model has been confirmed, nor is it entirely clear that the SoCQ yields reliable scores for all the model's constructs. Since the model is developmental, longitudinal studies

may be especially useful. It is also entirely possible that more sophisticated measurement models, such as generalizability theory (Eason, 1991), will yield important insights regarding this measurement protocol. In the meanwhile, the Stages of Concern Questionnaire might best be used with some caution.

References

Bailey, D.B., & Palsha, S.A. (1991). Measurement qualities of the Stages of Concern Questionnaire: Implications for documenting concerns about educational innovations. Manuscript submitted for publication.

Broyles, I., & Tillman, M. (1985). Relationships of inservice training components and changes in teacher concerns regarding innovations. Journal of Educational Research, 78, 364-371.

Cicchelli, T., & Baecher, R.E. (1987). The use of concerns theory in inservice training for computer education. Computer Education, 11, 85-93.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

Eason, S. (1991). Why generalizability theory yields better results than classical test theory: A primer with concrete examples. In B. Thompson (Ed.) Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 83-98). Greenwich, CT: JAI Press.

Hall, G.E. (1978). Teacher concerns as a basis for facilitating and personalizing staff development. Teacher College Reports, 80, 36-53.

Hall, G.E., George, A.A., & Rutherford, W.L. (1979/1986). Measuring stages of concern about the innovation: A manual for use of the SoC Questionnaire. Austin: Southwest Educational Development Laboratory.

Hall, G.E., & Hord, S.M. (1987). Change in schools: Facilitating

the process. Albany, NY: State University of New York Press.

Hall, G.E., & Loucks, S.F. (1981). Program definition and adaptation: Implications for inservice. Journal of Research and Development in Education, 14(2), 46-57.

Hall, G.E., Loucks, S.F., Rutherford, W.L., & Newlove, B. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. Journal of Teacher Education, 29, 52-56.

Holloway, W.H., & Kerr, M.E. (1979). A state-wide profile of the P.L. 94-142 related concerns of educators: Implications for planning and changing. Planning and Changing, 10, 246-256.

Hord, S.M., Rutherford, W., Huling-Austin, H., & Hall, G.E. (1987). Taking charge of change. Alexandria, VA: Association for Supervision and Curriculum Development.

James, R.K., & Hall, G.E. (1981). A study of the concerns of science teachers regarding an implementation of ISCS. Journal of Research in Science Teaching, 18, 479-487.

Kimpston, R.D., & Anderson, D.H. (1985). Teacher and principal concerns: The implementation of benchmark testing. Planning and Changing, 16, 96-104.

Locke, L.F., Spirduso, W.W., & Silverman, S.J. (1987). Proposals that work: A guide for planning dissertations and grant proposals (2nd ed.). Newbury Park, CA: SAGE.

Loucks, S.F., & Hall, G.E. (1977). Assisting and facilitating the implementation of innovations: A new approach. Educational Technology, 17, 18-21.

Neale, J.M., & Liebert, R.M. (1986). Science and behavior: An

introduction to methods of research (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Reinhardt, B. (1991, January). Factors affecting coefficient alpha: A mini Monte Carlo study. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio.

Rowley, G.L. (1976). The reliability of observational measures. American Educational Research Journal, 13, 51-59.

Sax, G. (1980). Principles of educational and psychological measurement and evaluation (2nd ed.). Belmont, CA: Wadsworth.

Wedman, J., & Heller, M. (1984). Concerns of teachers about educational computing. AEDS Journal, 18, 31-40.

Table 1
alpha Coefficients Across Studies and Scales

Stage/Scale	Sample Size 830[a]	142[b]	22[c]	15[c]	15[d]	25[e]	12[e]
0 Awareness	.64	.42	.12	**.29**	**.01**	.36	**-.56**
1 Informational	.78	.67	.48	**.54**	**.55**	.74	**.55**
2 Personal	.83	.77	.78	**.72**	**.22**	.80	**.70**
3 Management	.75	.64	.86	**.78**	**.74**	.62	**.57**
4 Consequence	.76	.79	.88	**.80**	**.72**	.67	**.73**
5 Collaboration	.82	.77	.61	**.70**	**.88**	.57	**.83**
6 Refocusing	.71	.61	.54	**.47**	**.19**	.80	**.39**

Note. Results from members of the longitudinal cohort are presented in **bold**.

[a]n for measurements reported by Hall et al. (1979).

[b]n for measurements reported by Bailey and Palsha (1991).

[c]n for measurements at Point 0, the pretest, in the present study.

[d]n for measurements at Point 1 in the present study.

[e]n for measurements at Point 2 in the present study.